DOCUMENT RESUME

ED 308 228 TM 013 601

AUTHOR Bridgeman, Brent
TITLE Comparative Validity of Multiple-Choice and Free-Response Items on the Advanced Placement Examination in Biology. College Board Report No. 89-2.
INSTITUTION College Entrance Examination Board, New York, N.Y.
REPORT NO ETS-RR-89-1
PUB DATE 89
NOTE 21p.
AVAILABLE FROM College Board Publications, Box 886, New York, NY 10101-0886 ($7.00).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Advanced Placement; *Biology; College Bound Students; College Entrance Examinations; *College Freshmen; Comparative Analysis; *Essay Tests; Grade Prediction; Higher Education; *Multiple Choice Tests; Objective Tests; Predictive Measurement; Predictor Variables; Sex Differences; *Test Validity
IDENTIFIERS *Advanced Placement Examinations (CEEB); Free Response Test Items; Placement Tests

ABSTRACT

The Advanced Placement Program reports grades to students and colleges on a 1 to 5 point scale derived from combining the separate scores on the multiple-choice and free-response sections of Advanced Placement Examinations (ADPEs). In this study, predictions of grades in first-semester college biology courses were made based on students' separate performances on the multiple-choice and essay portions of the 1986 ADPE in Biology. In a sample of 432 students from 11 colleges, the multiple-choice scores appeared to be more closely correlated to college grades. However, when the analyses were run separately according to the gender of the candidate, performance on both the objective and the essay items predicted grades equally well for males, but for females, predictions based on the essay portion were significantly less accurate. The current practice of reporting advanced placement examination grades based on a combination of both essay and multiple-choice test scores yields correlations that are generally higher and more comparable across gender than would be the case if only essays were used. Seven data tables and 14 figures plotting the students' scores are included. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED308228

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
[✓] This document has been reproduced as received from the person or organization originating it.
[ ] Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Robert G. Cameron

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

College Board Report

No. 89-2

Comparative Validity of Multiple-Choice and Free-Response Items on the Advanced Placement Examination in Biology

Brent Bridgeman

TM 013 601

Comparative Validity of Multiple-Choice and Free-Response Items on the Advanced Placement Examination in Biology

Brent Bridgeman

College Board Report No. 89-2
ETS RR No. 89-1

College Entrance Examination Board, New York, 1989

Brent Bridgeman is a senior research psychologist at Educational Testing Service, Princeton, New Jersey.

Acknowledgments

Thanks to Douglas Gentile for performing the data analyses and to Gloria Moreland for preparing the manuscript. Special thanks to the local data collectors at the participating institutions who spent many hours digging the necessary grade information out of departmental files.

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board is a nonprofit membership organization committed to maintaining academic standards and broadening access to higher education. Its more than 2,600 members include colleges and universities, secondary schools, university and school systems, and education associations and agencies. Representatives of the members elect the Board of Trustees and serve on committees and councils that advise the College Board on the guidance and placement, testing and assessment, and financial aid services it provides to students and educational institutions.

Additional copies of this report may be obtained from College Board Publications, Box 886, New York, New York 10101-0886. The price is $7.

Copyright © 1989 by College Entrance Examination Board. All rights reserved.
Advanced Placement Program, College Board, Scholastic Aptitude Test, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board.
Printed in the United States of America.

CONTENTS

Abstract 1

Introduction 1

Method 1
Test Description 1
Institution Selection 2
Procedures 2

Results 2
Gender Differences 5

Discussion 10

References 10

Tables

1. Numbers, Means, and Standard Deviations for Major Measures 3

2. Correlation of AP Biology Scores with College Biology Grades 4

3. Correlations of AP Biology Scores and Regular Grades with Laboratory Grades 5

4. Numbers, Means, and Standard Deviations According to Gender 8

5. Correlations of AP Scores and Course Grades by Gender 9

6. Correlations of AP Scores and Regular Grades with Laboratory Grades by Gender 9

Figures

1. Plot of AP Biology Essay Total Versus Objective Total for Subjects Having Valid Grades in Regular Biology Classes at College 3 5

2. Plot of AP Biology Essay Total Versus Objective Total for Subjects Having Valid Grades in Regular Biology Classes at College 4 6

3. Plot of AP Biology Essay Total Versus Objective Total for Subjects Having Valid Grades in Regular Biology Classes at College 9 6

4. Plot of AP Biology Essay Total Versus Objective Total for Subjects Having Valid Grades in Regular Biology Classes at College 10 7

Appendixes

A. College Descriptions 11

B. Test Scores and Grade Plots for Advanced Classes 12

C. Test Scores and Grade Plots by Gender 13

ABSTRACT

The Advanced Placement (AP) Program reports grades to students and colleges on a 1-to-5-point scale derived by combining the separate scores on the multiple-choice and free-response sections of AP Examinations.

Predictions of grades in first-semester college biology courses were made based on students' separate performances on the multiple-choice and essay portions of the Advanced Placement Examination in Biology. In a sample of 432 students from 11 colleges, the multiple-choice scores appeared to be more closely correlated to college grades. However, when analyses were run separately according to the gender of the candidate, performance on both the objective and the essay items predicted grades equally well for males (r [the correlation coefficient pooled across colleges] = .44 for the objective items and .46 for the essay section), whereas for females, predictions based on the essay portion were significantly less accurate (.49 versus .29). The current practice of reporting AP grades based on a combination of both essay and multiple-choice test scores yields correlations that are generally higher and more comparable across gender than would be the case if only essays were used.

INTRODUCTION

Advanced Placement (AP) Examinations are administered to secondary school students who have completed specific college-level courses. The examinations are used by colleges to grant credit for introductory-level work or to place students into more advanced courses, or both. The examinations typically include a set of multiple-choice questions and a set of free-response exercises consisting of essays or problem sets or both. Scores from the two parts are combined, usually with equal or slightly greater nominal weight on the free-response items, to form a single composite score. Cut points are established to convert the composite score to AP grades ranging from 1 to 5, which are the only scores reported to the colleges. Although the free-response items are relatively expensive to score, they assess productive skills that cannot be measured by multiple-choice questions, and they appear to measure skills seen as important in the college community. On the other hand, multiple-choice items efficiently use testing time to assess knowledge and concepts from a broad set of content areas.

Previous research comparing essay and objective test performance suggests that, as measures of writing ability, the two kinds of assessment are largely overlapping but that each also assesses some unique features (Breland and Gaynor, 1979; Quellmalz, Capell, and Chou, 1982). Both essay and objective tests are good predictors of performance in English composition courses, although the incremental validity of essays over objective scores appears to be rather small (Breland, Camp, Jones, Morris, and Rock, 1987). However, much less is known about the relative contributions of the two kinds of assessment techniques for measurement of subject-matter knowledge as opposed to assessment of writing ability.

The current study was designed to identify the contribution of each assessment technique as a predictor of success in college courses, as measured by grades. A finding of differential predictive validity might have implications for setting the relative weights of the two parts of the examination and would, in any event, provide test users with a more accurate understanding of the AP grade.

Biology was selected as the course area for this investigation because of the large number of students enrolled in AP Biology courses and the relative independence of the free-response and multiple-choice biology items on the examination. Note that although calculus and chemistry are also popular AP Programs, the available correlational evidence suggests that free-response and multiple-choice items on these examinations measure the same skills (Eignor, Flesher, and McLean, 1985a and 1985b); correlations between the two parts, corrected for unreliability, were .99 and 1.0 for calculus and chemistry, respectively. The correlation between the objective and free-response sections of the 1986 AP Examination in Biology that provided the data for the current study was .71 (between .80 and .87 corrected for unreliability). Thus the knowledge and skills measured by the two sections show a high degree of overlap, but each also appears to be measuring some unique components.

METHOD

Test description

The 1986 AP Examination in Biology consisted of a 90-minute objective section and a 75-minute free-response section. The objective section contained 120 five-choice items, 40 on each of three topics: cellular and molecular, organismal, and populational. The free-response (essay) section consisted of three pairs of questions, each covering one of the above topics. The candidate was directed to write on each topic by selecting one question from each pair.

The multiple-choice items were scored with a correction factor to compensate for random guessing. The essays were scored by a group of college and secondary school biology teachers brought together for this purpose. Each topic was graded on a 15-point scale; the

maximum score for the three essays was 45. Because each of the three essays by a particular candidate was read by a different reader, the kind of context effects discussed by Hales and Tokar (1975) should be minimized. The readers first conducted detailed discussions on scoring standards and then scored and discussed a set of sample papers. Reliability checks were made throughout the reading to make certain that judgments remained consistent among readers.

The lower limit for the reliability of the essay score, as estimated by coefficient alpha that included both readers and questions as sources of unreliability, was .73. The Kuder–Richardson Formula 20 (KR-20) reliability of the multiple-choice items was .93. A composite score with a maximum value of 150 was created by equally weighting the multiple-choice and essay sections (composite score = objective score × .625 + essay score × 1.667). After an equating process that ensures continuity of the score scale over time, the composite score was converted to the 1 to 5 grade reported to colleges (composite scores from 0 to 30 = 1, 31 to 47 = 2, 48 to 68 = 3, 69 to 91 = 4, 92 to 150 = 5).

Institution selection

Although the institutions invited to participate in this study did not form a true random sample, they did represent both public and private institutions from a range of geographical areas. A major selection criterion was that each institution had to be identified by substantial numbers of students as a college to which they wanted their AP Biology grades sent. (Because AP testing takes place in May, after most students have made college choice decisions, it was hoped that most students would, in fact, attend the college to which they had asked their grades be sent.) Because AP candidates tend to be very good students, this selection criterion meant that most institutions were relatively selective.

Out of 20 institutions initially contacted, 11 sent usable data. Some of the institutions that declined to participate did so because of departmental policies that discouraged students, with or without AP credits, from taking biology during the freshman year. Only courses taken during the freshman year were considered in this study because it was felt that a year of intervening coursework would unduly dilute the correlations of AP grades with course grades. See Appendix A for a brief description of each institution in the sample.

Procedures

Paid local coordinators at each institution were sent lists of students who had taken the 1986 AP Examination in Biology and asked that their scores be sent to that institution. For each student on the list, the local coordinator recorded the grade received in the first biology course taken during the freshman year. Where available, separate grades for lecture and laboratory sections were recorded. The coordinator classified the courses into one of four categories: (1) the introductory course taken by most freshmen, (2) a different freshman-level course (including courses in botany and zoology), (3) an introductory course specifically designated for students not majoring in biology, or (4) an advanced course open only to students with advanced placement. Grades were coded on a 13-point scale (F = 0, D− = 1, D = 2, . . . , A+ = 12).

RESULTS

Sample sizes in each kind of course at each institution, along with means and standard deviations on the major measures, are presented in Table 1. Because within-course correlation coefficients were to be conducted, only courses that contained at least four students who had taken the AP Examination were included in the table and in the analyses. Only 10 subjects were eliminated by this criterion.

Given the relatively large number of students who sent their AP grades to each institution (see Appendix A), the small numbers in each course were quite surprising. Out of a total of 1,277 names on the initial rosters, usable data were returned for only 432 subjects. Follow-up investigation revealed that the local data coordinators did a good job of tracking students, but that a number of factors contributed to the low proportion of the original sample that could provide grade information. First, even though the decision as to where to send AP grades was not made until May of the senior year, many students did not attend the institution to which they sent their scores. For example, out of 71 students who sent grades to College 7, 23 did not enroll in the college. Second, and more important, a large number of students did not take biology in the freshman year, either because of institutional policy or because they chose to take other courses. As mentioned above, some colleges declined the invitation to participate because they preferred students to take their first biology course in the second year, and some colleges that agreed to participate had similar policies even though they allowed a few freshmen to take biology courses. Even in the absence of an institutional policy, many students apparently elected to defer further study of biology until later in their college careers or not to take any biology courses in college, perhaps because many had received a full year of college credit in biology through the AP Program. Except for the small minority of students majoring in the life sciences, most majors would not require additional biology courses.

Table 1. Numbers, Means, and Standard Deviations for Major Measures

			AP Objective		*AP Essay*		*AP Composite*		*AP 1–5*		*Course Grade*	
	College	*N*	*Mean*	*(SD)*	*Mean*	*(SD)*	*Mean*	*(SD)*	*Mean*	*(SD)*	*Mean*	*(SD)*
Regular	1	10	60.3	(9.6)	18.2	(4.8)	68.2	(12.1)	3.30	(.48)	8.8	(1.9)
introductory	2	4	78.5	(9.3)	19.5	(1.7)	81.8	(7.3)	4.00	(.00)	7.0	(2.4)
course	3	85	71.2	(14.3)	17.9	(6.4)	74.4	(16.9)	3.68	(.80)	7.0	(2.3)
	4	38	59.6	(14.1)	15.4	(5.7)	63.0	(16.1)	3.16	(.72)	5.8	(2.3)
	5	10	54.9	(18.8)	11.1	(8.2)	52.9	(22.3)	2.60	(1.17)	6.6	(4.1)
	6	8	37.0	(13.4)	7.3	(2.8)	35.0	(12.6)	1.63	(.74)	5.9	(3.0)
	7	17	49.8	(19.6)	11.5	(5.1)	50.3	(17.4)	2.41	(.94)	5.5	(3.0)
	8	13	36.9	(16.1)	8.2	(4.6)	36.7	(14.4)	1.92	(.76)	7.0	(3.3)
	9	38	52.0	(16.7)	13.3	(7.3)	54.7	(21.2)	2.76	(1.08)	7.8	(2.8)
	10	30	53.3	(17.0)	12.7	(6.6)	54.7	(18.3)	2.77	(1.04)	6.3	(4.0)
Total		253										
Other	1	18	62.2	(12.1)	17.3	(11.0)	67.7	(23.1)	3.44	(.92)	8.3	(2.4)
introductory	3	11	78.0	(13.0)	19.8	(8.5)	81.8	(21.0)	3.91	(.83)	6.9	(3.0)
course	4	25	59.9	(14.3)	14.1	(5.3)	61.0	(14.0)	3.12	(.73)	7.2	(2.9)
	5	14	56.8	(16.4)	15.9	(8.4)	61.9	(23.5)	3.00	(1.04)	5.9	(3.9)
	9	4	42.8	(17.6)	12.0	(7.5)	46.8	(20.6)	2.25	(1.26)	6.5	(3.9)
	10	14	66.8	(14.2)	15.7	(6.8)	67.9	(17.8)	3.43	(.85)	8.4	(1.6)
Total		86										
Nonmajors	2	7	47.3	(12.9)	10.0	(6.5)	46.3	(18.0)	2.14	(.90)	8.0	(2.4)
course	3	8	57.8	(2.7)	14.6	(2.2)	60.4	(5.1)	3.00	(.00)	7.0	(3.0)
	8	5	46.6	(30.0)	12.4	(15.1)	50.0	(42.0)	2.20	(1.64)	9.4	(3.6)
	10	6	46.5	(18.6)	8.5	(2.6)	43.2	(15.3)	2.00	(.89)	8.5	(2.3)
	11	5	86.6	(14.7)	22.6	(9.3)	91.8	(24.2)	4.40	(.89)	8.0	(2.5)
Total		31										
Advanced	3	23	88.7	(11.9)	28.2	(7.9)	102.5	(16.4)	4.83	(.39)	8.1	(2.6)
course	5	8	70.4	(9.9)	20.0	(5.8)	77.5	(14.4)	4.00	(.53)	7.6	(3.0)
	6	13	70.2	(15.5)	19.0	(7.5)	75.5	(19.5)	3.85	(.99)	5.8	(3.1)
	7	7	80.9	(19.8)	23.9	(6.8)	90.3	(19.5)	4.29	(.76)	6.7	(2.4)
	9	5	76.8	(11.9)	21.8	(5.6)	84.4	(11.3)	4.40	(.55)	10.4	(1.3)
	10	6	86.2	(17.5)	28.0	(7.9)	100.7	(23.3)	4.50	(.84)	8.2	(4.3)
Total		62										

Another surprising finding revealed in Table 1 is the number of students enrolled in introductory-level courses, even among students with high AP grades. Out of the 432 students in the sample, only 62 were in advanced courses. Again the explanation appears to be a combination of institutional policies and personal preferences. For example, at College 3 (which had the largest number of students in the sample) students not majoring in biology had to have at least an AP grade of 4 to receive credit; students expecting to take advanced biology courses had to have a grade of 5 in order to skip the introductory course, and even then taking the introductory course was recommended if the student was in doubt. The chairman of the biology department at another institution reported that many students who could have been exempted from the introductory course decided to take it in the mistaken belief that it would be an easy A that would improve chances of entry to medical school or other graduate study.

Table 1 also reveals the expected variability in AP performance levels across institutions and the necessity for calculating separate correlations for each kind of course within an institution. Note, for example, that the AP composite score means were 20 points (or about .8 standard deviations based on the national norm data) higher for students in the regular introductory course at College 3 than for students in the comparable course at College 9, yet the course grade means were actually slightly higher at College 9. Within one institution (College 6), AP composite score means were 40 points higher in the advanced course than in the regular introductory course, but the course grade means were nearly

identical (5.8 compared to 5.9). If all subjects from a single kind of course were combined across institutions, or if subjects from different courses within an institution were considered as a single group, the resulting correlations would be nearly meaningless.

Table 2 presents the correlations between each of the AP scores and college grades, as well as the weighted average of the correlations across institutions for each kind of course and the total weighted average across institutions and kinds of courses. The weighted average was computed by multiplying the correlation by the numbers (N) on which it was based, adding these products, and dividing by the total *number*. Because of the very small numbers at some institutions, not too much should be made of the fluctuations from one institution to the next. However, considerable confidence can be placed in the overall conclusion, based on 432 students, that the objective score is superior to the essay score as a predictor of grades in the first biology course taken in college. Also note that the composite score is no better as a predictor than the objective score by itself.

Separate grades for the laboratory component of regular introductory courses were available from three institutions. Correlations of the AP scores with these laboratory grades, as well as the correlation of regular grades with laboratory grades, are presented in Table 3.

The predictive superiority of the objective scores is again apparent. College 3 reported that the laboratory grade was based on laboratory reports and short-

Table 2. Correlation of AP Biology Scores with College Biology Grades

			*AP Biology Scores**			
	College	*N*	*Objective*	*Essay*	*Composite*	*1–5*
Regular introductory course	1	10	52	27	44	19
	2	4	98	39	97	00
	3	85	51	18	38	37
	4	38	32	37	39	27
	5	10	78	58	77	79
	6	8	−23	−08	−17	−28
	7	17	81	38	76	84
	8	13	65	42	68	70
	9	38	37	45	44	47
	10	30	43	63	63	64
Weighted total		253	47	34	47	43
Other introductory course	1	18	38	35	40	51
	3	11	−01	28	19	32
	4	25	49	35	53	34
	5	14	55	66	62	58
	9	4	96	72	94	92
	10	14	51	58	62	70
Weighted total		86	44	44	51	50
Nonmajors course	2	7	34	28	31	15
	3	8	25	−31	−15	00
	8	5	49	24	35	41
	10	6	15	36	21	00
	11	5	75	46	57	66
Weighted total		31	37	17	22	23
Advanced course	3	23	33	−05	11	−20
	5	8	57	25	42	54
	6	13	29	65	55	58
	7	7	78	23	63	52
	9	5	60	−42	07	41
	10	6	95	81	91	93
Weighted total		62	48	22	37	30
Weighted grand total		432	46	33	45	41

*Decimal points have been omitted.

Table 3. Correlations of AP Biology Scores and Regular Grades with Laboratory Grades

			AP Scores				*Regular*
	College	*N*	*Objective*	*Essay*	*Composite*	*1–5*	*Grade*
	3	84	39	20	33	33	60
	8	13	66	23	59	66	62
	9	36	36	40	41	39	38
Weighted total		133	41	26	38	38	54

answer quizzes. Note, however, that at College 9 the essay score was as accurate as the objective score in predicting both the laboratory grade and the regular grade (see Table 2). There is insufficient data available in the current study to pinpoint the reasons for this apparent discrepancy, but it should be noted that regular grades at both College 3 (where the discrepancy between essay and objective correlations was large) and College 9 (where the discrepancy was small and in the opposite direction) were based on multiple-choice tests.

A correlation coefficient is a good index of general group performance, but it can sometimes mask information about the performance of atypical individuals. For example, one might want to know about the performance of students who wrote superior essays but were only average on the objective section (or vice versa). With the sample sizes in the current study, it is possible to get a direct pictorial representation of these relationships. For each of the four largest regular introductory courses, grades were plotted in the space defined by the objective and essay scores (see Figures 1–4). Although median splits and study of the outliers did not exhibit any consistent trends, the plots revealed some interesting patterns. At three out of the four colleges, the student with the highest essay score and the student with the highest objective score each received a B grade. College 9 demonstrates that correlation is not destiny and gives some hope to students with low AP grades; the student with the lowest essay score (about a 3) and the student with the lowest objective score (about a 17) each received an A in the course.

Appendix B provides similar information for the two largest advanced classes. Again, no particularly striking pattern emerged, although it is interesting to note that the four students who earned C's and the one who earned a D at College 3 were near the middle of both distributions, whereas some students who scored lower on both the essay and objective sections earned A's.

Gender differences

Nationally the AP Examination in Biology is taken by about as many males as females (in 1986 the figures were 12,936 males and 12,842 females). On the

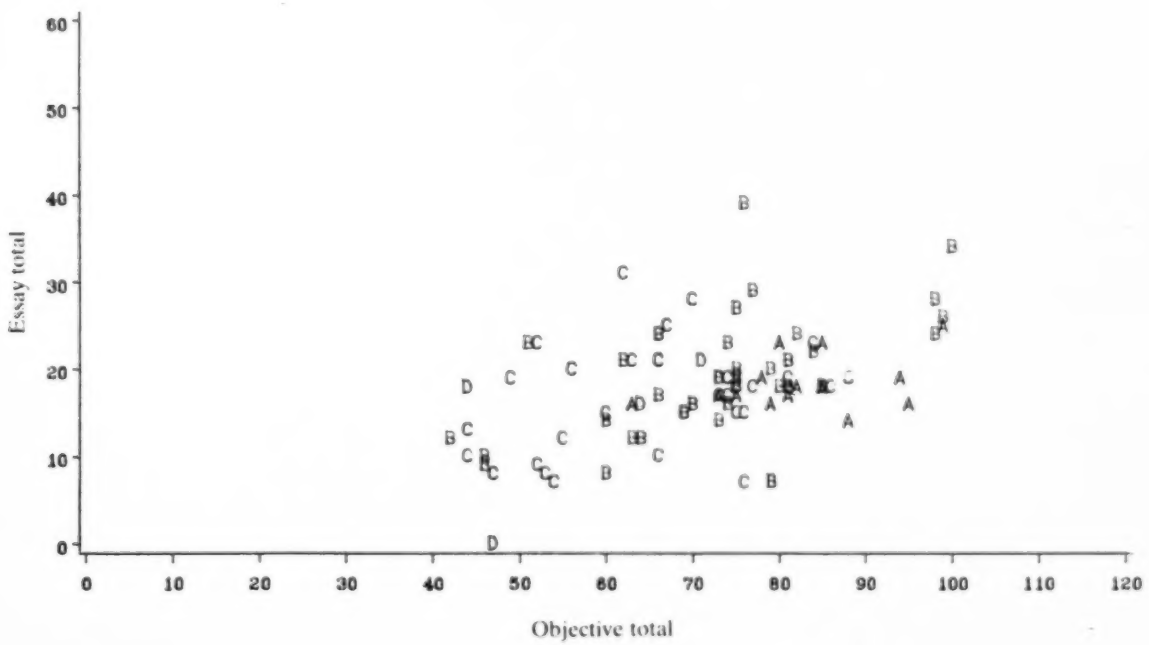


Figure 1. Plot of AP Biology Essay Total Versus Objective Total for Subjects Having Valid Grades in Regular Biology Classes at College 3 (number = 85)

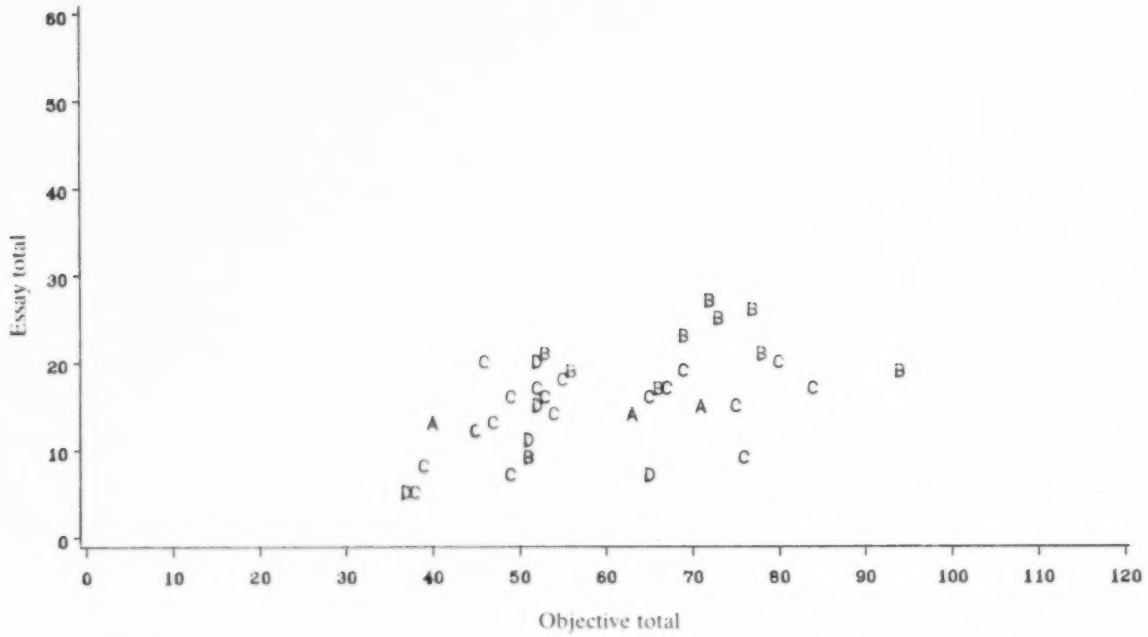


Figure 2. Plot of AP Biology Essay Total Versus Objective Total for Subjects Having Valid Grades in Regular Biology Classes at College 4 (number = 38)

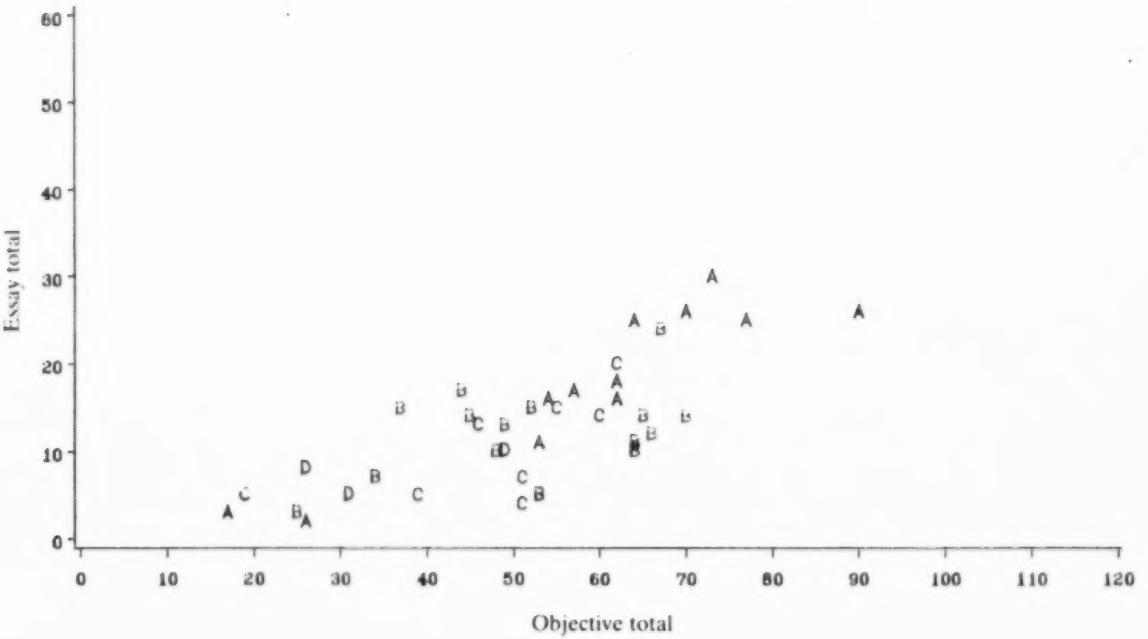


Figure 3. Plot of AP Biology Essay Total Versus Objective Total for Subjects Having Valid Grades in Regular Biology Classes at College 9 (number = 38)

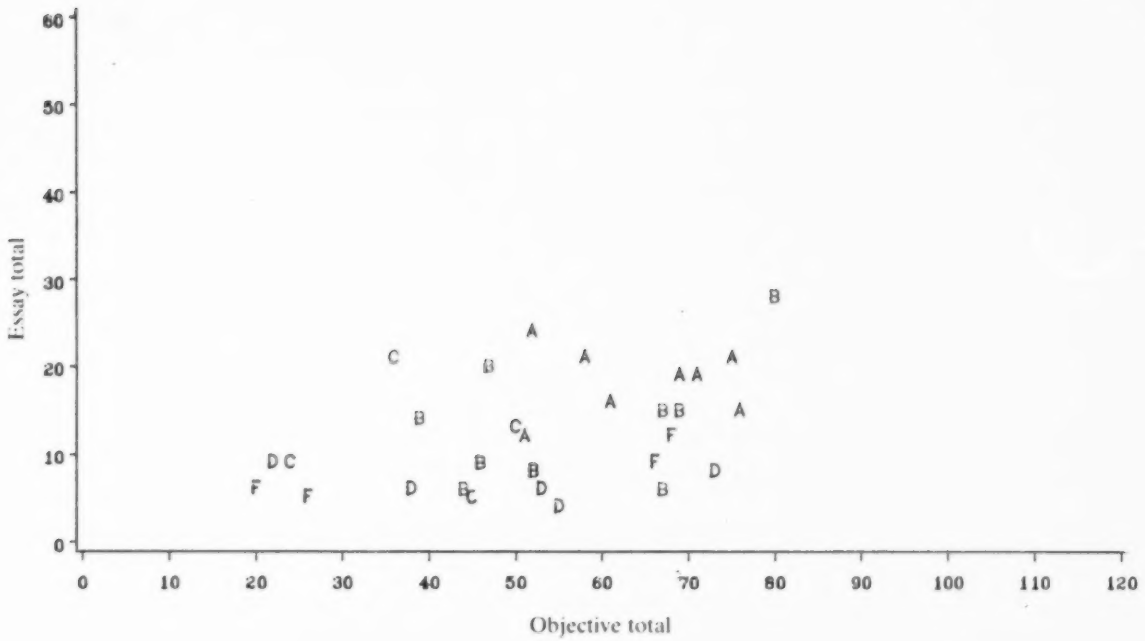


Figure 4. Plot of AP Biology Essay Total Versus Objective Total for Subjects Having Valid Grades in Regular Biology Classes at College 10 (number = 30)

multiple-choice items, score means for males were about one-third of a standard deviation higher than scores for females (62.8 versus 55.9; $SD = 20.5$ in both groups). On the essay items the difference was only about one-tenth of a standard deviation (16.1 versus 15.1; $SD = 8.8$ in both groups). Table 4 presents comparable figures for each course in the current sample in which there were at least four students of each sex. Also included in Table 4 are the grade means for each course according to gender.

Consistent with the national data, note that the gender difference in the objective scores is generally greater than the difference in the essay scores. There is also a hint that the test, as a predictor of college grades in the regular introductory courses, might be biased in favor of females (or that college grading might be biased against females). In only one of the seven courses were grade means for females higher than those for males, even though AP essay scores for females were higher in five of the seven courses and AP objective scores for females were higher in four of the courses. Note especially that in the relatively large samples at College 4 and College 10, grades for females were about one-half of a standard deviation lower than grades for males, even though females scored higher on both parts of the AP Examination. But even in the relatively large samples, the trend was not totally consistent; at College 9, grade differences were consistent with score differences. In the "other introductory courses" the general direction of the bias appeared to be reversed, with females receiving higher grades than males despite having lower test scores. However, interpretations of grades in this category are more problematic because of the small sample sizes (the largest male sample is nine students) and the variety of different courses (such as botany and zoology) included under the "other" umbrella.

Table 5 presents the correlational results separately by gender. It is apparent that the overall group findings presented in Table 2 reflect a difference that occurs only among females. The weighted grand total correlations indicate that objective and essay scores were equally good predictors of college grades for males, whereas for females essays were significantly poorer predictors, $t(198) = 2.95$, $p \leq .01$. Median correlations computed over the 15 categories in Table 5 were remarkably similar to the weighted grand total averages. Also note that in 11 of the 15 samples of female students, scores on the objective items correlated more closely to grades than did scores on the essay items; this pattern was noted in only 5 of the 15 samples of male students. Plots of course grades in the space defined by the essay and objective AP scores, separately by gender, are presented in Appendix C.

Combining the correlations reported under "weighted grand totals" in a multiple regression analysis yielded nearly equal standardized regression weights on the two test sections for male students (.31 for the essay

Table 4. Numbers, Means, and Standard Deviations According to Gender

				AP Biology Scores								*Course Grade*	
				Objective		*Essay*		*Composite*		*1–5*			
	College	*Sex*	*N*	*Mean*	*(SD)*	*Mean*	*(SD)*	*Mean*	*(SD)*	*Mean*	*(SD)*	*Mean*	*(SD)*
Regular	1	M	4	57.0	(13.9)	18.8	(5.0)	67.0	(15.1)	3.25	(.50)	9.5	(1.7)
introductory		F	6	62.5	(6.1)	17.8	(5.2)	69.0	(11.1)	3.33	(.52)	8.3	(2.1)
course	3	M	34	74.1	(15.9)	17.4	(7.0)	75.3	(20.2)	3.68	(.98)	7.5	(2.3)
		F	51	69.2	(12.9)	18.3	(6.0)	73.8	(14.5)	3.69	(.68)	6.8	(2.3)
	4	M	18	58.2	(15.0)	14.6	(6.0)	60.6	(18.6)	3.06	(.87)	6.4	(2.1)
		F	20	60.9	(12.3)	16.3	(5.4)	65.1	(13.5)	3.25	(.55)	5.3	(2.3)
	7	M	10	57.9	(14.6)	10.8	(5.0)	54.3	(15.3)	2.70	(.82)	6.5	(3.2)
		F	7	38.1	(20.9)	12.4	(5.5)	44.6	(19.7)	2.00	(1.00)	4.1	(2.3)
	8	M	5	35.0	(19.8)	8.0	(5.3)	35.2	(18.8)	1.80	(.84)	5.4	(3.3)
		F	8	38.1	(14.6)	8.3	(4.5)	37.6	(12.3)	2.00	(.76)	8.0	(3.1)
	9	M	15	55.3	(16.2)	14.4	(6.7)	58.5	(19.7)	3.00	(.93)	8.0	(3.0)
		F	23	49.9	(17.1)	12.6	(7.7)	52.3	(22.3)	2.61	(1.16)	7.7	(2.7)
	10	M	15	51.8	(15.8)	11.6	(5.5)	51.9	(16.9)	2.53	(.92)	7.5	(3.3)
		F	15	54.9	(18.6)	13.8	(7.5)	57.5	(19.8)	3.00	(1.13)	5.2	(4.4)
Other	1	M	4	66.3	(7.4)	19.0	(14.6)	73.3	(24.7)	3.75	(.96)	7.8	(2.6)
introductory		F	14	61.1	(13.2)	16.1	(10.4)	66.1	(23.3)	3.36	(.93)	8.4	(2.4)
course	3	M	6	81.0	(15.5)	20.2	(8.7)	63.0	(1.4)	3.80	(.84)	4.0	(5.7)
		F	5	74.4	(9.8)	19.4	(9.3)	79.0	(21.2)			7.8	(0.8)
	4	M	8	66.0	(13.7)	13.8	(5.1)	64.4	(12.0)	3.25	(.71)	7.1	(3.8)
		F	17	57.0	(14.0)	14.3	(5.5)	59.5	(14.9)	3.06	(.75)	7.2	(2.6)
	5	M	9	59.4	(17.1)	16.3	(8.8)	64.4	(24.4)	3.00	(1.00)	5.4	(3.9)
		F	5	52.0	(15.6)	15.0	(8.5)	57.4	(23.6)	3.00	(1.22)	6.8	(4.0)
	10	M	6	66.8	(13.1)	13.7	(6.5)	64.5	(15.1)	3.17	(.41)	7.5	(1.2)
		F	8	66.8	(15.9)	17.3	(7.0)	70.5	(20.3)	3.63	(1.06)	9.1	(1.6)
Advanced	3	M	12	91.8	(12.9)	27.3	(6.6)	102.8	(15.1)	4.83	(.39)	8.4	(2.5)
course		F	11	85.2	(10.1)	29.3	(9.2)	102.1	(18.1)	4.82	(.40)	7.8	(2.8)
	5	M	4	70.0	(13.4)	21.8	(8.2)	80.3	(21.1)	4.00	(.82)	7.3	(2.9)
		F	4	70.8	(6.9)	18.3	(1.7)	74.8	(3.8)	4.00	(.00)	8.0	(3.5)
	6	M	6	76.3	(13.1)	22.2	(4.7)	84.7	(18.3)	4.17	(.98)	6.8	(3.9)
		F	7	64.9	(16.3)	16.3	(6.9)	67.7	(17.9)	3.57	(.98)	4.9	(2.2)

portion and .26 for objective items). But for female students the standardized regression weights were .45 for the objective section and only .12 for the essays. The multiple correlations were .50 and .51 for males and females, respectively.

Table 6 indicates that the findings for differential prediction of laboratory grades for essay and objective sections reported in Table 3 were also solely a function of a difference in the female sample, $t(78) = 3.27$, $p \leq .01$.

Evidence from the AP Examination itself suggests that essay performance is less closely related to the skills measured by the multiple-choice items for females than for males. As indicated in Table 5, weighted grand totals for essay and objective items were correlated .59 for males but only .39 for females, $Z = 2.46$, $p \leq .01$ (after r to Z transformation). Again, median correlations echoed the results found using the weighted averages. However, when the data for all 201 females in the sample were combined, without respect to college and course designation, the correlation between objective and essay items jumped to .60, and the comparable figure for the 156 male subjects went up only to .70; the sex difference became nonsignificant (*ns*) $t(355) = 1.55$.

In the national sample of all candidates taking the AP Examination in Biology (as estimated from a spaced sample of 3,216), the correlation between essay and objective sections was .70 for males and .73 for females. It is not surprising that the correlations in the relatively homogeneous course and college categories would be smaller than in the heterogeneous total study sample or in the national sample, but the emergence of the gender difference is more difficult to explain. One hypothesis is that if females were more sensitive to course demands they might do a better job of selecting the most appropriate course, resulting in a lower within-course variance (compared to males) and hence a lower within-course correlation. However, this hypothesis is not supported by the data. Note, for example, in Table 4 that for the regular course at College 10, performance by females was more variable than by males, but the correlation was still greater for males. Overall, the weighted-average standard deviations of the composite scores were nearly identical for males and females (17.7 versus 17.1).

Table 5. Correlations of AP Scores and Course Grades by Gender

	College	Sex	N	AP Scores with Course Grades: Objective	Essay	Composite	1–5	AP Essay with AP Objective
Regular introductory course	1	M	4	78	17	54	58	64
		F	6	73	29	49	06	48
	3	M	34	40	36	40	43	76
		F	51	58	07	37	33	28
	4	M	18	27	35	34	18	74
		F	20	46	51	60	53	30
	7	M	10	76	47	71	81	51
		F	7	89	56	85	88	55
	8	M	5	46	66	61	76	54
		F	8	86	30	81	68	10
	9	M	15	69	70	75	77	71
		F	23	16	30	25	32	80
	10	M	15	63	71	76	79	58
		F	15	39	73	69	75	34
Weighted totals		M	101	51	48	54	55	68
		F	130	50	30	48	45	39
Other introductory course	1	M	4	−53	22	11	36	−02
		F	14	55	42	51	59	62
	3	M	6	03	34	23	44	79
		F	5	56	47	49	29	93
	4	M	8	68	58	90	73	−02
		F	17	43	22	39	12	39
	5	M	9	46	62	57	48	85
		F	5	93	82	87	76	96
	10	M	6	03	58	44	20	20
		F	8	87	49	72	81	74
Weighted totals		M	33	24	50	51	47	40
		F	49	60	41	54	45	49
Advanced course	3	M	12	36	−05	16	−11	30
		F	11	25	−03	06	−29	29
	5	M	4	47	71	65	85	85
		F	4	88	−85	38	00	−50
	6	M	6	68	78	81	79	61
		F	7	−37	36	03	20	38
Weighted totals		M	22	47	31	42	31	48
		F	22	17	−06	11	−08	18
Weighted grand totals		M	156	44	46	52	50	59
		F	201	49	29	45	39	39
Median*		M		46	47	54	58	58
		F		55	30	40	33	38

*Computed over 15 course and college categories.

Table 6. Correlations of AP Scores and Regular Grades with Laboratory Grades by Gender

	College	Sex	N	AP Scores: Objective	Essay	Composite	1–5	Regular Grades
	3	M	33	31	41	39	44	55
		F	51	50	03	30	24	66
	8	M	5	81	64	84	89	80
		F	8	60	01	46	53	44
	9	M	14	50	66	65	72	70
		F	22	34	33	35	32	23
Weighted total		M	52	41	50	50	56	61
		F	81	47	11	33	29	52

DISCUSSION

If prediction of freshman biology grades were the primary purpose of the Advanced Placement Examination in Biology one might reasonably conclude that the essay items should be eliminated because of the observed sex differences; predictions for females could be made as well or better based on the objective items alone. However, because content validity is of paramount importance for such knowledge certification tests as the AP Examinations, mere lack of predictive validity or gender differences in predictive validity is not an adequate reason to restructure the examination. This is particularly true given the nature of the criterion: grades in large freshman courses are determined almost exclusively by scores on multiple-choice tests. It should not be surprising that one set of multiple-choice items predicts another set of similar items. Even the short-answer quizzes used to compute a major portion of the laboratory grade at College 3 may contain cognitive demands that are more similar to multiple-choice tests than they are to the organized, extended discourse required in the AP essays.

If the ability to write sophisticated essays on selected topics in biology is seen as an important skill, it has a place in the AP Examination that is unaffected by gender differences in predictive efficiency. Furthermore, the gender difference might disappear if the grades used as the criterion were determined more by essay-test results than by multiple-choice or short-answer examinations. Nevertheless, the results of this study may be considered as support for the current practice of using both multiple-choice and essay scores to compute the 1 to 5 AP grade. If only essay results were used to compute the grade, predictions of college-course grades would be substantially less accurate, especially for female students.

Why would the correlation between the essay and objective sections of the AP Examination show a gender difference at the individual course level but not when the subjects are aggregated without regard to course distinctions? There may be some complex differential self-selection by males and females into courses, but further research with large samples is needed for a full understanding of the factors involved.

REFERENCES

Breland, H.M., R. Camp, R.J. Jones, M.M. Morris, and D.A. Rock. 1987. *Assessing Writing Skills*. New York: College Entrance Examination Board.

Breland, H.M. and J. Gaynor. 1979. A Comparison of Direct and Indirect Assessment of Writing Skill. *Journal of Educational Measurement* 16: 119–128.

Eignor, D.R., R.B. Flesher, and D. McLean. 1985a. *Test Analysis: College Board Advanced Placement Examination in Chemistry*. Princeton, NJ: Educational Testing Service. Unpublished Statistical Report No. SR-85-172.

Eignor, D.R., R.B. Flesher, and D. McLean. 1985b. *Test Analysis of College Board Advanced Placement Examinations—Mathematics: Calculus AB, Mathematics: Calculus BC*. Princeton, NJ: Educational Testing Service. Unpublished Statistical Report No. SR-85-149.

Hales, L.W. and E. Tokar. 1975. The Effect of the Quality of Preceding Responses on the Grades Assigned to Subsequent Responses to an Essay Question. *Journal of Educational Measurement* 12:115–117.

Quellmalz, E.S., F.J. Capell, and C. Chou. 1982. Effects of Discourse and Response Mode on the Measurement of Writing Competence. *Journal of Educational Measurement* 19:241–258.

APPENDIX A. COLLEGE DESCRIPTIONS

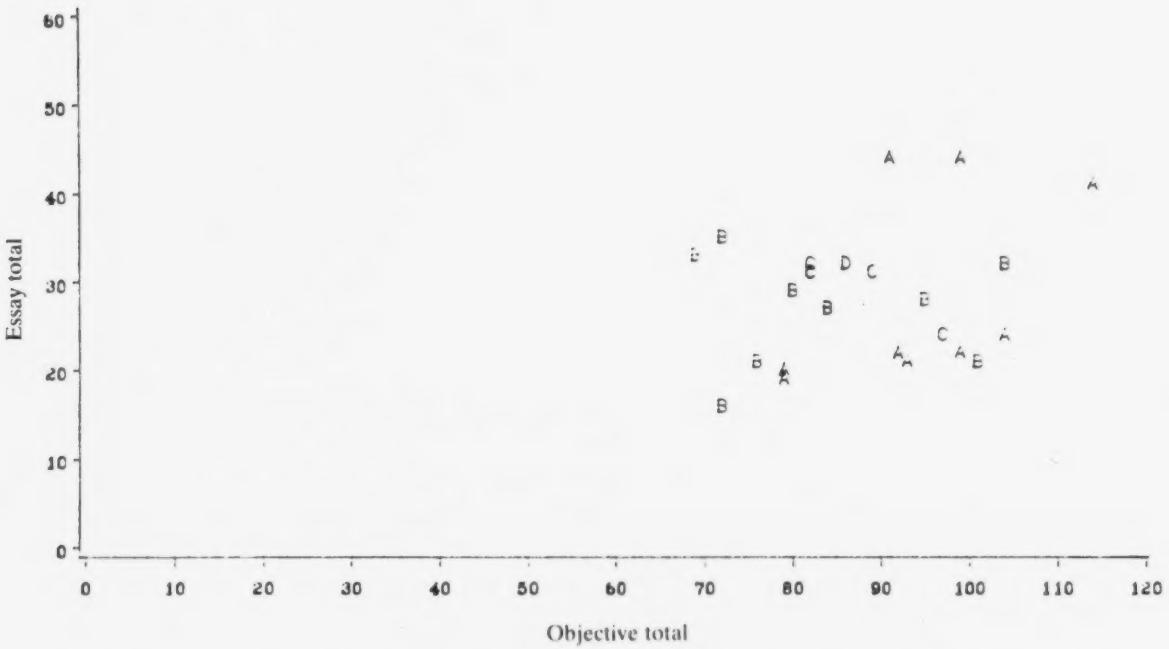
College	*Enrollment**	*AP Credit***	*Scores Sent†*
1. Midwestern state university	18,000	3	270
2. Midwestern church-affiliated university	7,000	5	63
3. Eastern private university	12,000	4 (5 for majors)	265
4. Eastern state university	7,000	3	152
5. Rocky Mountain state university	14,000	3	53
6. Rocky Mountain state university	13,000	3	87
7. Southern private university	8,000	3	71
8. Southern state university	14,000	3	75
9. Southern state university	28,000	4	79
10. Southwestern state university	37,000	3	119
11. Southwestern private university	3,000	4	43

*Undergraduate enrollment figures are given to the nearest thousand.

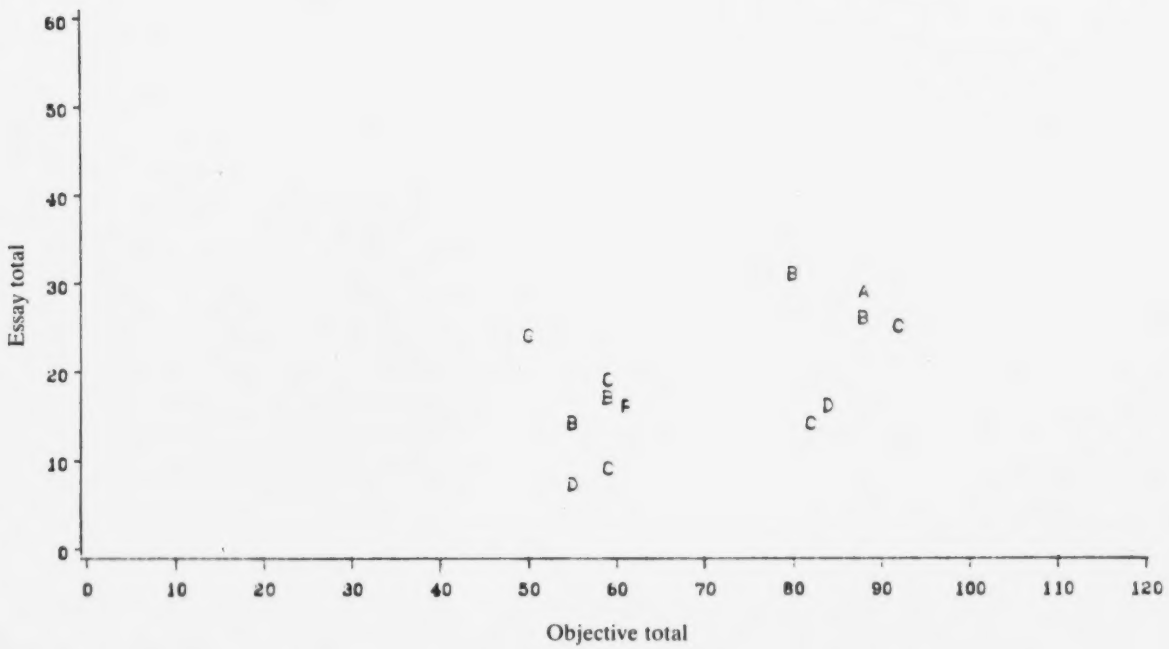
**AP credit represents the minimum grade on the AP Examination in Biology for which credit is granted at the institution.

†Figures represent the number of students who requested that their grades on the AP Examination in Biology be sent to the institution.

APPENDIX B. TEST SCORES AND GRADE PLOTS FOR ADVANCED CLASSES

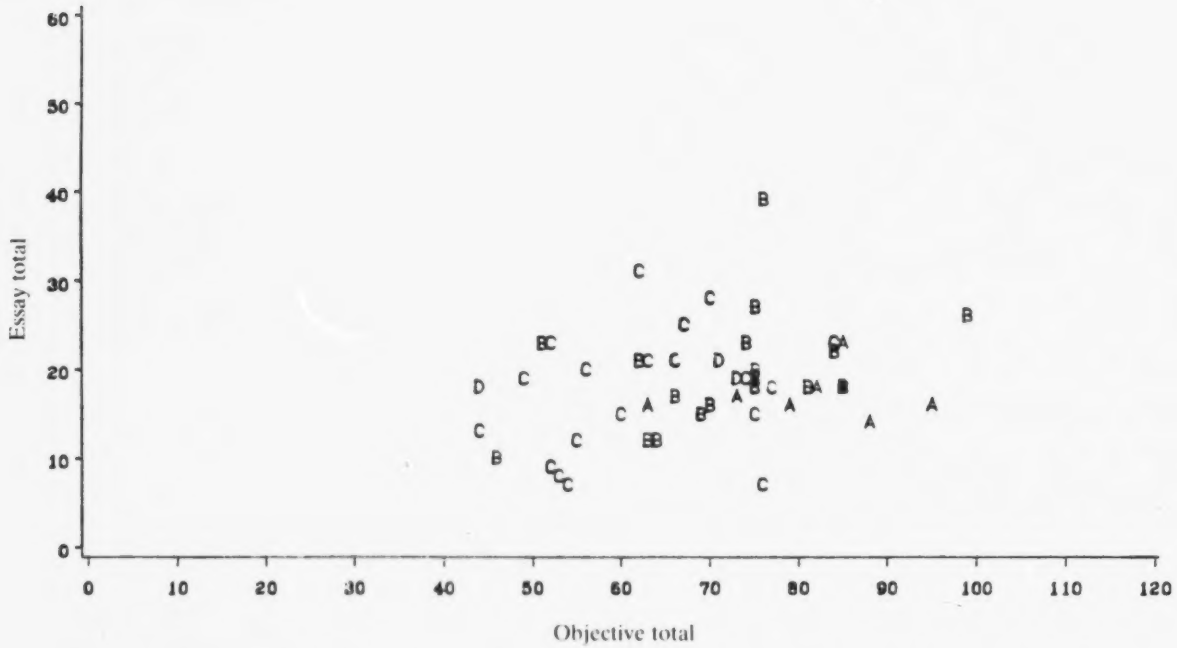


Plot of AP Biology Essay Total Versus Objective Total for Subjects with Valid Grades in Advanced Biology Classes at College 3 (number = 23)

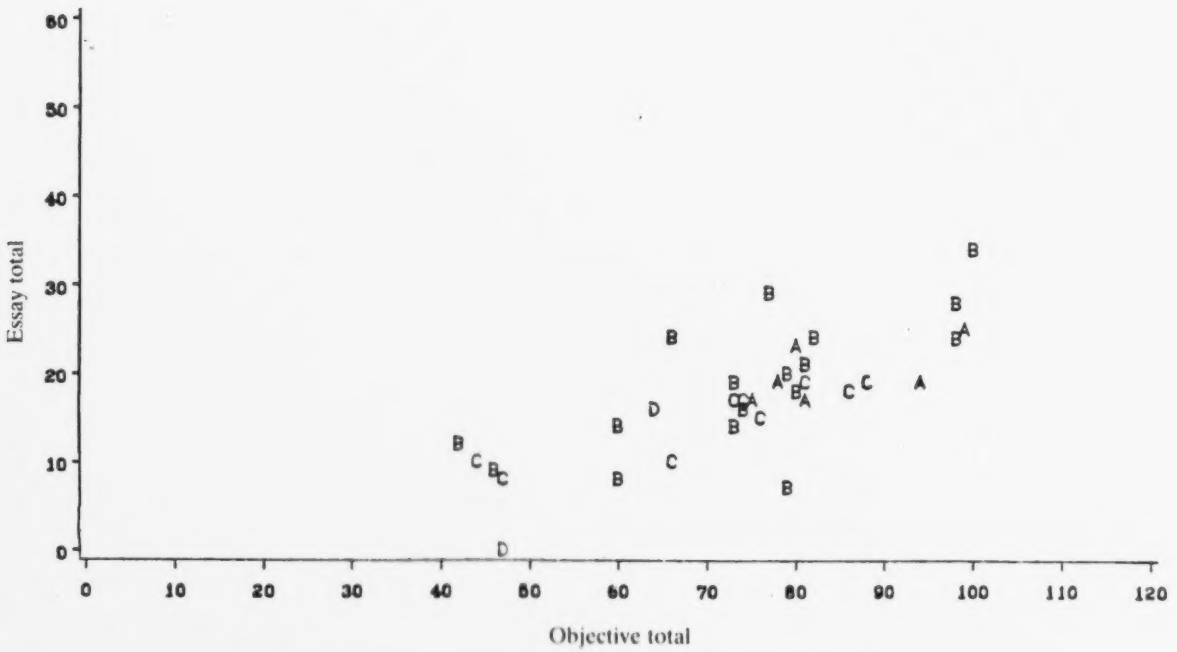


Plot of AP Biology Essay Total Versus Objective Total for Subjects with Valid Grades in Advanced Biology Classes at College 6 (number = 13)

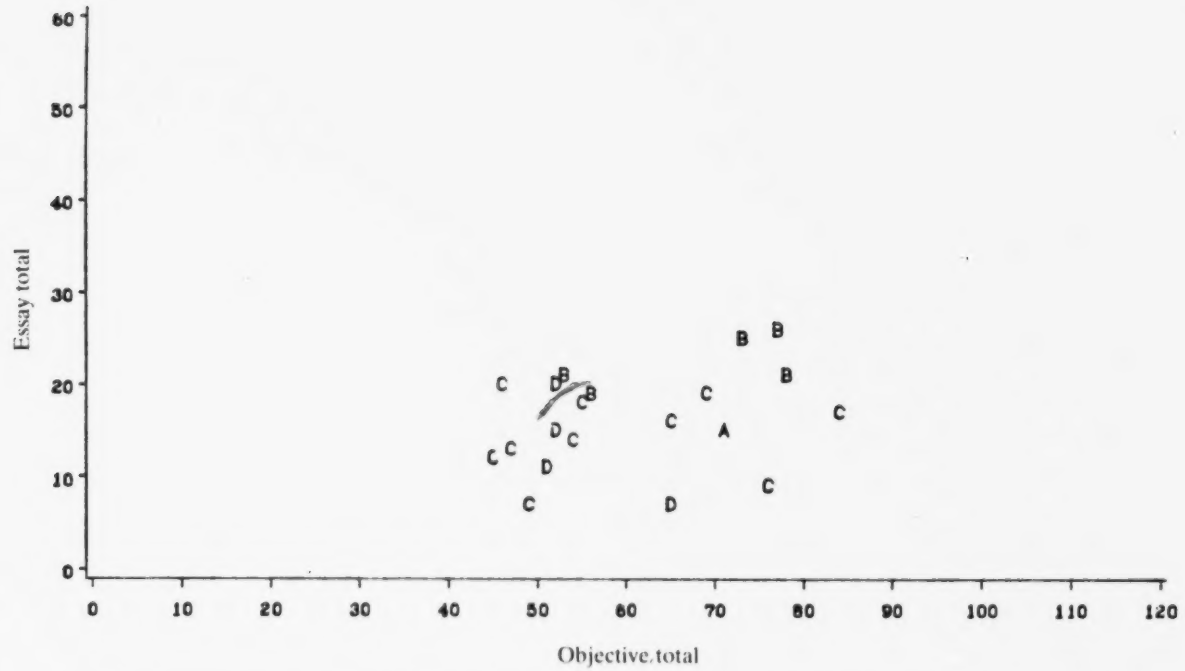
APPENDIX C. TEST SCORES AND GRADE PLOTS BY GENDER



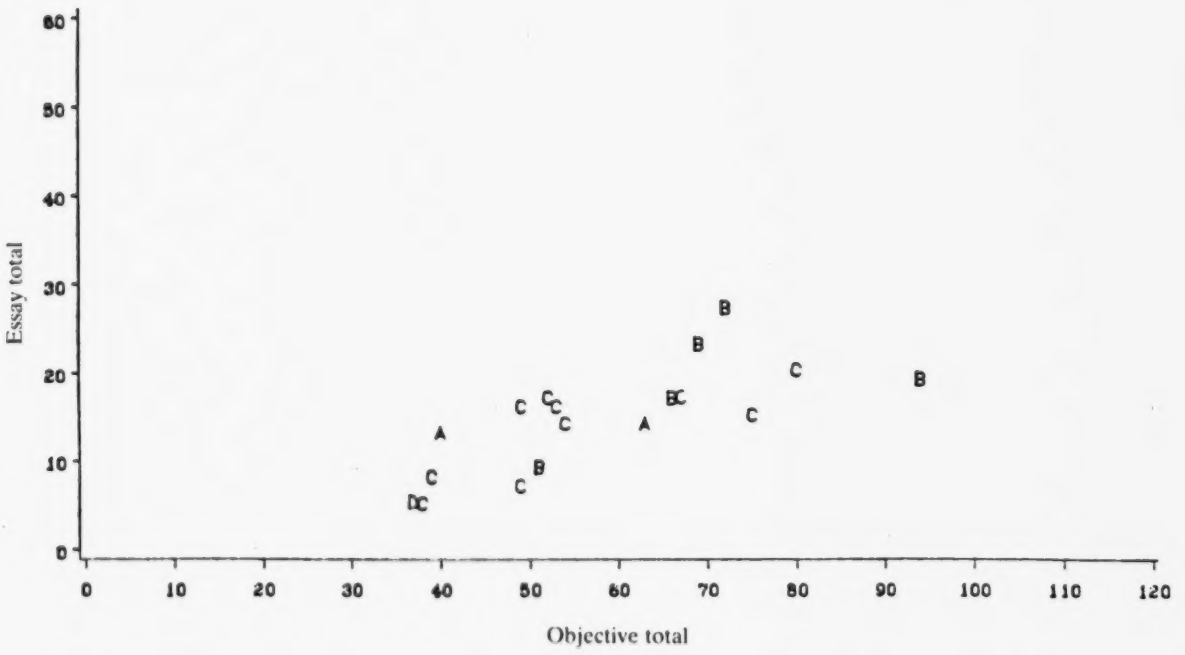
Plot of AP Biology Essay Total Versus Objective Total for Females with Valid Grades in Regular Biology Classes at College 3 (number = 51)



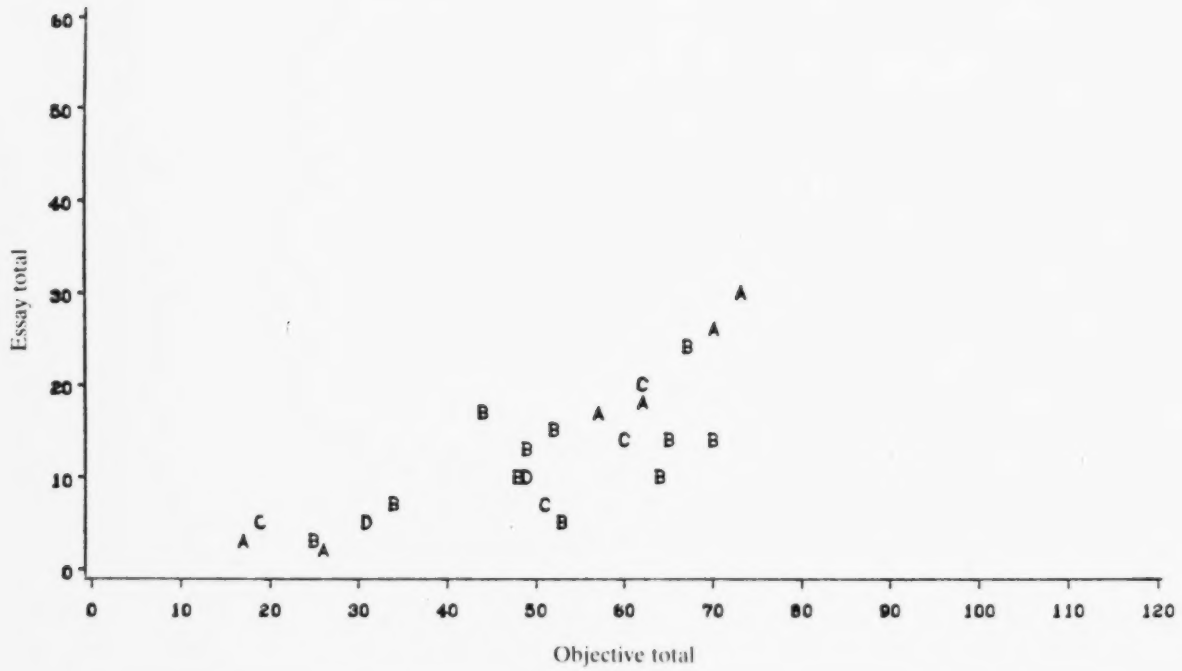
Plot of AP Biology Essay Total Versus Objective Total for Males with Valid Grades in Regular Biology Classes at College 3 (number = 34)



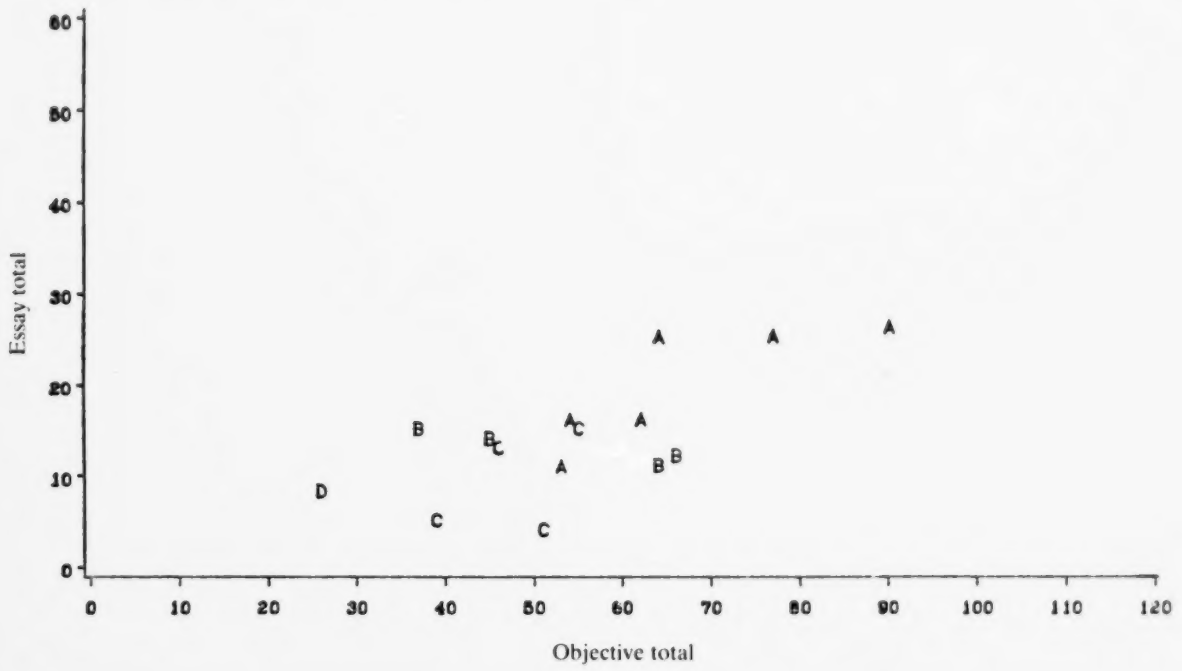
Plot of AP Biology Essay Total Versus Objective Total for Females with Valid Grades in Regular Biology Classes at College 4 (number = 20)



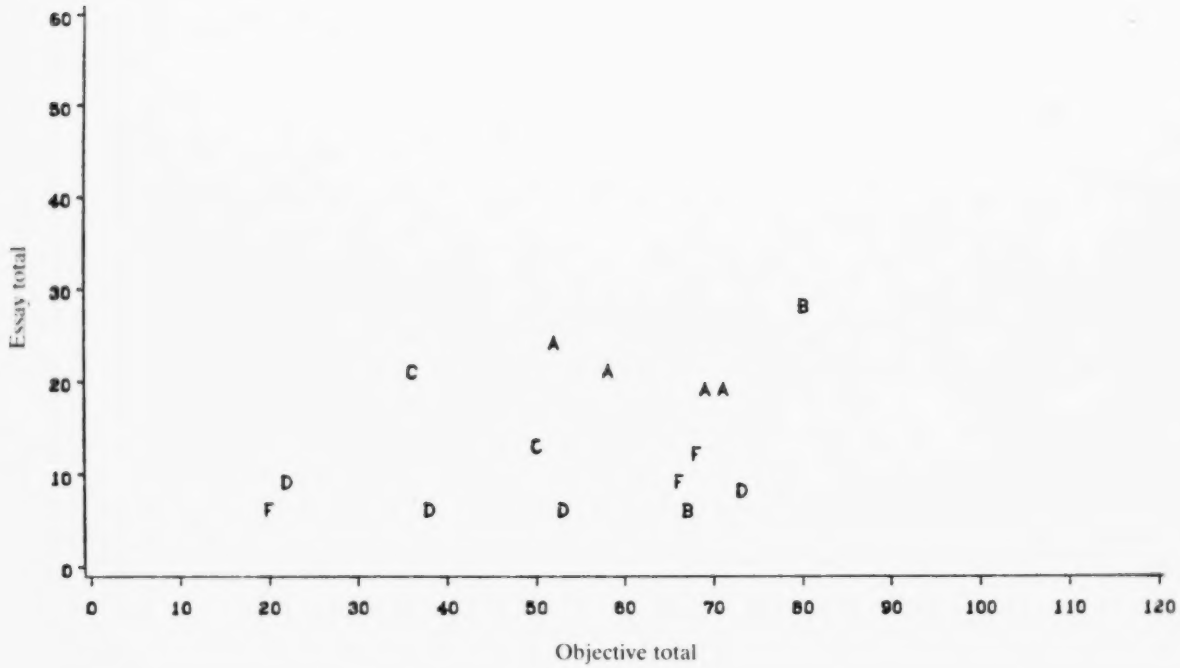
Plot of AP Biology Essay Total Versus Objective Total for Males with Valid Grades in Regular Biology Classes at College 4 (number = 18)



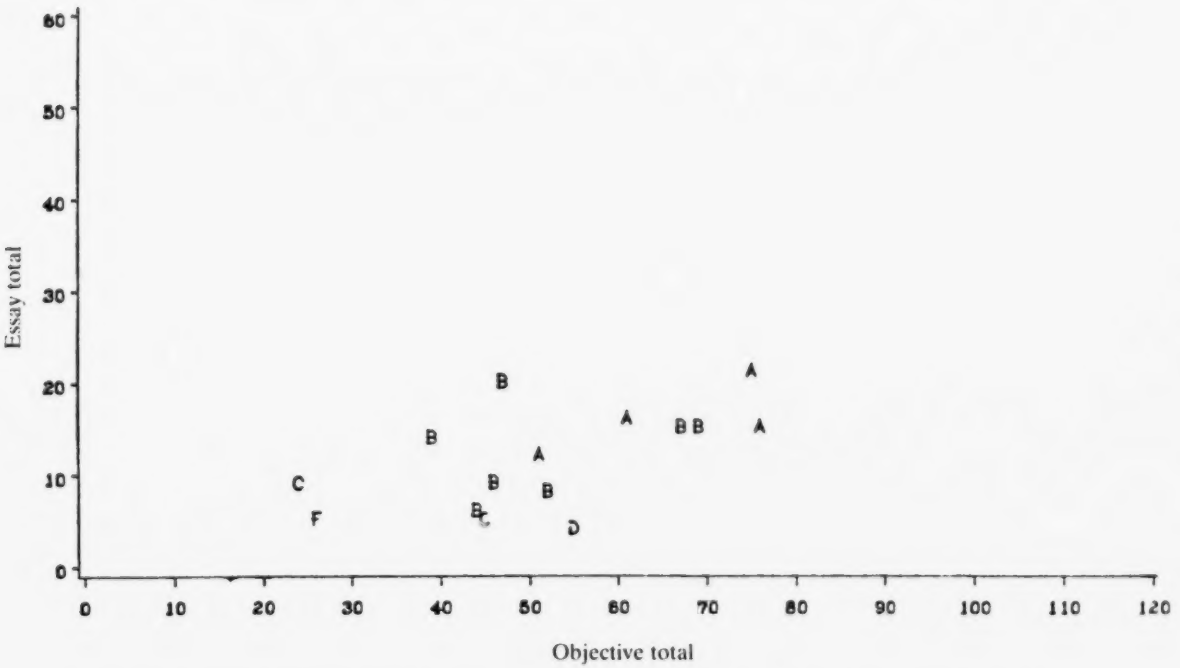
Plot of AP Biology Essay Total Versus Objective Total for Females with Valid Grades in Regular Biology Classes at College 9 (number = 23)



Plot of AP Biology Essay Total Versus Objective Total for Males with Valid Grades in Regular Biology Classes at College 9 (number = 15)



Plot of AP Biology Essay Total Versus Objective Total for Females with Valid Grades in Regular Biology Classes at College 10 (number = 15)



Plot of AP Biology Essay Total Versus Objective Total for Males with Valid Grades in Regular Biology Classes at College 10 (number = 15)